Actions, not gestures: contextualising embodied controller interactions in immersive virtual reality

Jack Ratcliffe
Queen Mary University of London
London, UK
j.ratcliffe@qmul.ac.uk

Nicholas Ballou
Queen Mary University of London
London, UK
n.b.ballou@qmul.ac.uk

Laurissa Tokarchuk
Queen Mary University of London
London, UK
laurissa.tokarchuk@qmul.ac.uk

ABSTRACT

Modern immersive virtual reality (IVR) often uses embodied controllers for interacting with virtual objects. However, it is not clear how we should conceptualise these interactions. They could be considered either gestures, as there is no interaction with a physical object; or as actions, given that there is object manipulation, even if it is virtual. This distinction is important, as literature has shown that in the physical world, action-enabled and gesture-enabled learning produce distinct cognitive outcomes. This study attempts to understand whether sensorimotor-embodied interactions with objects in IVR can cognitively be considered as actions or gestures. It does this by comparing verb-learning outcomes between two conditions: (1) where participants move the controllers without touching virtual objects (gesture condition); and (2) where participants move the controllers and manipulate virtual objects (action condition). We found that (1) users can have cognitively distinct outcomes in IVR based on whether the interactions are actions or gestures, with actions providing stronger memorisation outcomes; and (2) embodied controller actions in IVR behave more similarly to physical world actions in terms of verb memorization benefits.

CCS CONCEPTS

• **Human-centered computing** → **Virtual reality**; **HCI theory, concepts and models**; **Gestural input**; • **Applied computing** → **Interactive learning environments**.

KEYWORDS

immersive virtual reality, sensorimotor, cognition, HCI, embodiment, virtual reality, learning

ACM Reference Format:
Jack Ratcliffe, Nicholas Ballou, and Laurissa Tokarchuk. 2021. Actions, not gestures: contextualising embodied controller interactions in immersive virtual reality. In *27th ACM Symposium on Virtual Reality Software and Technology (VRST '21), December 8–10, 2021, Osaka, Japan.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3489849.3489892

VRST '21, December 8–10, 2021, Osaka, Japan

ACM ISBN 978-1-4503-9092-7/21/12…$15.00
https://doi.org/10.1145/3489849.3489892

1 INTRODUCTION

Modern consumer immersive virtual reality (IVR) is increasingly leveraging sensorimotor-embodied controllers as their predominant input method (e.g. controllers such as the Oculus Touch, Vive Wands or Index Knuckles). However, despite their increasing ubiquity in both consumer software and research studies, there has been little research into how we cognitively contextualise these controller-mediated interactions within IVR.

This ambiguity is reflected in literature discussing embodied controllers, in which they are also referred to as gesture controllers or hand gesture inputs [21][23]. These systems are also sometimes referred to as natural user interfaces [40], despite the use of embodied controllers being quite *un*natural. For example, in order to act-out drinking from a virtual cup in IVR using the controllers listed above, a user must find an open space in the physical world, grasp a plastic controller in a grip similar to how you would hold a gun or TV remote, and move until a virtual presentation of the hand reaches the virtual cup, then bring the virtual cup to their virtual avatar's head position and await system feedback that the drinking action occurred. These movements are depicted in Fig. 1.

This is fairly distinct from the action of drinking from an actual cup in the physical world, and could be categorised as a gesture, given that our physical bodies move in an abstracted way and do not interact physically with the target object. Equally, it could be categorised as an action, as we physically act on the controllers; or, if we examine the virtual space, our physical movements allow us to virtually act on virtual objects. This distinction is important, as whether a movement is an action or a gesture has consequences for learning outcomes enabled by different embodied cognition approaches.

Actions, defined as movements on or using objects, have a different cognitive framework and present evidence of different cognitive outcomes than gestures, defined as movements about objects. Learning with actions, generally, has been shown to make stronger and more specific mnemonic impressions on people experiencing them or enacting them, whether that is for the location of objects [14], or the memorisation of words [58]. They have also been found to be easier for learners to process [16].

Alternatively, learning with gestures has been shown to promote better representational rather than absolute understanding of objects [35], and an enhanced ability to generalise verbs to wider situations [34][58].

In order to understand if a similar distinction between action and gesture exists in IVR, we propose investigating differences in learning outcomes between groups memorising verbs that have been encoded with either action-based or gesture-based interaction. Any distinctions between these two conditions would suggest that

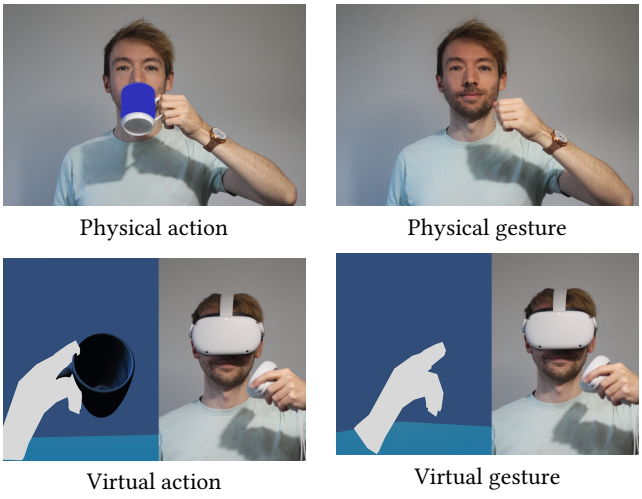


Figure 1: Images showing actions and gestures in physical and virtual world

embodied controller interactions in IVR are able to be conceptualised as physical world actions, while a lack of distinctions would support the idea that embodied IVR interactions should only be conceptualised in a way similar to physical world gestures.

We created an IVR system for learning Japanese verbs in which half the participants learned via actions (being able to manipulate verb-congruent virtual objects) and half learned via gestures (not being able manipulate virtual objects). If there was no difference in learning outcome between the groups, then that would provide evidence that embodied controller interactions could be considered as gestures and could not be considered as actions, as the actions did not provide the additional memorisation benefits suggested in literature.

However, our results found that there were different learning outcomes between the conditions. This suggests that embodied controllers allow us to have cognitively distinct experiences in IVR, and that IVR inputs are not just "gestures", but depend on how the interaction is designed in the IVR environment.

As the action group provided better learning benefits, in the same way as in physical world studies, these results also suggest that we could cognitively respond to actions in IVR in a similar way to physical world actions, and to gestures in IVR in a similar way to real-world gestures.

These findings presents two implications for IVR language-learning software design: (1) actions are more effective than gestures at promoting verb encoding; and (2) the outcomes similarly map to the physical world in terms of sensorimotor verb encoding. Beyond these, the more generalized implication for embodied interface design is that the choice of action-based or gesture-based embodied interaction in VR system designs can have notable and distinct impacts on the user's cognitive outcomes.

2 LITERATURE

2.1 Sensorimotor embodiment benefits; action and gesture distinctions

Sensorimotor activity is generally considered to have an impact on cognition, particularly learning. Many studies demonstrate evidence that learners memorise information better when they encode it while performing congruent sensorimotor activities [5][37]. This is known as the enactment effect or self-performed task (SPT) effect [6][44][4]. Experiments have shown that both taking actions with objects (SPT-Os) [12] and gesturing without objects aid the memorisation process.

Recent research has suggested that there could be learning distinctions between two types of sensorimotor activity: encoding with actions (e.g. kicking a ball) and gestures (e.g. just kicking) [58]. Actions, defined as movements on or using objects, present evidence of different cognitive outcomes than gestures, defined as movements about objects [14].

Learning with actions, generally, has been shown to make stronger and more specific mnemonic impressions on people experiencing them or enacting them, whether that is for the location of objects [14], or the memorisation of words [57]. They have also been found to be easier for learners to process [16].

Learning with gestures has been shown to promote better representational rather than absolute understanding of objects [35], and an enhanced ability to generalise verbs to wider situations [34][58].

In comparative studies between actions and gestures, Wakefield and Hall [57] found that children learned novel verbs better through action experiences rather than gesture experiences (although they later found similar rates of learning [58]). There have also been higher rates of recognition and recall accuracy for verbs with a greater amount of associated information [50].

There have been numerous explanations for the learning distinctions between action-based and gesture-based learning. The first is that acting-on-objects is cognitively distinct from gesturing-off-objects, and uses different encoding routes, even if the movements are similar [58]. Evidence for this exists in the distinction between physical manipulation theories [30] and gesture-simulated action [13] approaches to embodied cognition.

The second explanation is that the distinction can be explained by the enactment increasing the distinctiveness of the memory traces by adding item-specific and relational information [36]. If we make an action on an object (in the physical or virtual worlds), and the object reacts, we experience the object as manipulable, and there is evidence that the perceived manipulability of an object impacts how we memorise it [27]. In this, Madan and Singhal interpreted the overall benefit for highly manipulatable items as being due to automatic activation of motor representations. Perhaps actions-on-objects stimulate these to a higher degree then gestures-off-objects?

A third explanation is that the enactment effect is not caused be sensorimotor encoding, but by the enhanced "learning episode" the sensorimotor activity creates [18]. By enacting an action like "lifting the pen", the act of lifting and the pen are registered together in a single episode. This view suggests that actions-on-objects creates deeper episodic integrations than gestures-off-objects. Supporting this view is evidence that semantically sensible learning situations cause stronger memorisation outcomes. For example, Mangels and

Heinberg found that semantically sensible action phrases (e.g. "hug the doll") had better memorisation outcomes then stranger ones (e.g. "hug the shovel"), suggesting semantic association played a role in memorisation [29]. Relating this to actions and gestures, perhaps taking actions-on-objects creates a more semantically sensible learning situation than gesturing-off-objects, and hence the noted learning effect.

2.2 Sensorimotor IVR benefits; action and gesture distinctions?

There is growing evidence that encoding information using sensorimotor activity inside IVR can also provide learning benefits over non-sensorimotor alternatives. Research into IVR second language learning has shown greater efficacy in sensorimotor scenarios than in non-sensorimotor ones. Vasquez found that verbs were remembered better if encoded by performing congruent actions than if learning using traditional text-based memorisation [56]. Ratcliffe found that a combination of verbs and nouns were remembered better if they were encoded by performing actions with objects in IVR, than if there were those actions were not performed but the objects were still present [38]. Fuhrman found improved learning rates for nouns that were learned when using a relevant sensorimotor activity compared to an irrelevant sensorimotor activity or no sensorimotor activity [7]; while Macedonia found similar [25].

However, whether there are distinctions in cognitive and learning outcomes in IVR between action-on-object encoding and gesture-off-object encoding is under-explored. We illustrate the distinction between actions and gestures in an IVR system in Fig. 2, with both gestures and actions requiring a user's bodily activation, but with actions also requiring virtual objects to be manipulable or for the environmental to give feedback in response to the bodily activation.

Although there is little experimental evidence for a cognitive distinction between IVR actions and gestures, there is some evidence that IVR actions are similar to their physical world counterparts. Studies into sensorimotor IVR skill development have shown that improvements transfer from virtual to physical world domains [9][19][54]. A neuromuscular investigation into throwing in the real-world and (non-immersive) virtual reality using electromyography signals of 11 muscles of the upper limbs also found a very high similarity between the virtual and physical actions [46]. However, another study found that throwing precision and accuracy in IVR are lower, and that it requires more user effort and produces a different kinematic throwing pattern [59].

Similarly, an argument that IVR actions are experienced in way similar to physical world actions could be made based on investigations into the IVR body transfer illusion [51]. According to this research, users perceive the actions of other agents on their virtual bodies in a similar way to real actions, rather than actions happening to a distinct avatar. However, that does not mean that the inverse is true: that actions a user takes in IVR are considered a physical actions rather than gestures.

From a mediated-interactionist perspective, the boundary between a cognitive agent and his or her environment can be considered malleable [1], and so it follows that IVR actions that have

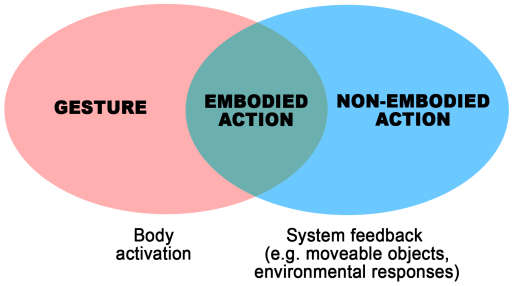


Figure 2: Diagram showing proposed distinctions between gestures and action in IVR. Actions (both embodied and non-) feature two system-created feedback points: interactional (user is able to move objects) and environmental (the world responds to the user's movement of objects). The overlap of the system feedback and body activation is embodied action, which in this paper we compare with gesture

similar outcomes to those of physical world actions may encourage our brains to think we are taking actions in IVR, and not just outputting gestures.

2.3 Other learning-related IVR factors

Sensorimotor activity can affect other cognitive factors, such as presence and motivation (see [39] for a summary). These factors can then affect learning outcomes. Both higher senses of spatial or perceptive presence (the feeling of being in a place) and motivation/engagement are linked to enhanced learning outcomes [33][45][20], and adding sensorimotor interaction, both in IVR and outside of it, increases users' feelings of both motivation [22][53][11][24][49] and presence [2][32][47].

In IVR learning research it is unclear if the use of sensorimotor interactions has a mediating or direct effect on learning, however, there is evidence that sensorimotor interactions should be treated as a direct contributor to IVR learning [39], similar to the physical world sensorimotor learning studies mentioned above.

Attempts to understand the impact and experience of sensorimotor activity and embodiment in IVRs are emerging, such as to a quantify users' sense of embodiment in IVR through an embodiment questionnaire [8]. This approach breaks embodiment into six sub-scales for differing experimental interests, of which two forms of embodiment are relevant to this research - agency and body ownership.

Further work on embodiment in IVR has also shown that task performance differs based upon the virtual avatar the user is embodying [17][3][10], with experiments showing participants being more expressive or performing better cognitively based upon embodying particular avatars. It is possible that, as has been theorised with motivation and presence, avatar type could have an interaction effect with sensorimotor activity.

2.4 Summary

The literature presents evidence for distinct cognitive outcomes for action-encoded and gesture-encoded learning in the physical world,

particularly in language learning. It also presents some evidence that for skill-based learning, actions taken in IVR are similar to those in the physical world.

However, there remains a lack of exploration into learning distinctions resulting from gesture-encoding or action-encoding in IVR sensorimotor learning, and whether these are similar to those in the physical world. Therefore investigating whether there is a distinction in language learning outcome between the encoding types in IVR could further our understanding of types of sensorimotor embodiment in IVR, and their relationship with physical world sensorimotor activities.

3 EXPERIMENT

We ran a between-subject experiment to investigate if there was a distinction between encoding with actions or gestures on verb memorisation in IVR. For the action condition, participants were able to use objects to complete actions in order to learn a congruent verb. For example, a participant had to pick up a cup, bring it to their mouth and tilt it while learning the Japanese word for "drink". In the gesture condition, participants had to make a gesture relevant to the verb, but were unable to interact with the object (i.e. could not touch or move the cup). In both conditions, an animated 3D model demonstrated the action/gesture for the verb.

We monitored and compared the learning gain of each condition to understand the role that interactions play in cognitive and memorisation of verbs. We also monitored participant embodiment, presence and motivation scores to investigate potential correlations between these metrics and our findings.

3.1 Hypotheses

Our hypotheses are based on literature that presents actions as more powerful verb encoders than gesture. We also present hypotheses that distinctions between actions and gestures might reflect in affective factors related to IVR, such as embodiment, presence and motivation results:

- h1. The action group will demonstrate stronger verb learning gains than the gesture group
- h2. The action group will demonstrate faster response times than the gesture group
- h3. The action group will report stronger embodiment than the gesture group
- h4. The action group will report stronger presence than the gesture group
- h5. The action group will report stronger motivation than the gesture group

3.2 Procedure

Participants were asked to download and run an executable file on their existing IVR systems. An on-boarding process, pre-test, learning process and post-test all took place within the downloaded software.

The on-boarding process explained the IVR control methods, and required users to move to a target location to continue the experience. A voice-over explained the experiment goals and process. The on-boarding process also gave an interactive tutorial of how the learning process worked before launching a pre-test.

Participants were pre-tested for their knowledge of 15 Japanese verbs. The pre-test involved listening to a Japanese word and choosing its English meaning from a list of 15 verbs, or skipping the question. Questions were presented sequentially and participants were not allowed to amend previous answers.

Participants were assigned to one of two interaction groups: action or gesture. They differ as follows:

- Action: Input is made by grabbing the actual VR object, and doing the correct gesture with it. A complete action is given some kind of feedback (e.g. drinking sounds for acting out a drinking motion)
- Gesture: Input is made by doing the correct gesture in the air, away from the object

During the experiment, participants were asked to memorise 15 verbs (see Table 1 for list). Participants were exposed to each verb in sequence, for five sequences. Each verb related to a different object presented in front of the participant on a podium in the IVR. Participants were told an action/gesture, the English verb, and the Japanese language verb. For example, a phrase used for learning "drink" was "drink from the cup. Drink is nomimasu. Nomimasu. Nomimasu". A 3D animation of a human doing each gesture was also displayed.

For each verb, the participant had to either gesture or action once (depending on their group) and say the verb aloud once. We instructed participants to say the verb aloud in order to control for the Production Effect, in which speaking a word while encoding it causes stronger memorisation than not speaking it [26]. Both groups of participants used the same avatar - a set of white hands with no arms or body.

After the encoding process, participants repeated the pre-test procedure. Learning gain was calculated as the final test result minus the pre-test results. A further, web-browser-based test was taken one week after the initial study to determine their retention of the information.

The data collection was done inside a VR environment. To verify the gestures and actions were completed correctly, telemetry of the participant's movements was recorded.

3.3 Participants

Fifty-six (56) participants took part of in our study. Of these, 53 were compensated and three were uncompensated. Uncompensated participants volunteered to take part after compensation offers had closed and this change had been advertised.

Forty-eight (48) participants' data was usable in our analysis. Two participants were excluded for having high levels of pre-existing Japanese knowledge (they already knew six and eight of the 15 target words). One participant was excluded for presenting unusual movement data. A follow-up conversation revealed they were using a spoofed virtual reality system (i.e. they used a monitor, mouse, keyboard and emulator to access and play VR content). Five participants were excluded due to incomplete data being returned from the remote software, and not manually forwarding the data when requested.

All valid participants who reported their recruitment referrer (38 participants) came from an advertisement posted on the Reddit /r/oculus community and used their own IVR hardware in their

Table 1: List of target words, the user's actions for encoding (for the action condition) and the feedback given by the system when an action was successfully completed

Verb	Action	System feedback
Wear	Pick-up a hat and place on head	Hat sticks to head, appears at top of vision
Wash	Pick-up a plate and place in a sink	Plate submerges in water, washing sound plays
Drink	Pick-up a cup, bring to mouth and tilt	Drinking sound plays
Smoke	Pick-up a cigarette and bring to mouth	Inhaling and exhaling sound places
Climb	Place hands on vertical climbing rope	Player is raised into the air as if climbing
Open	Pick-up a a box lid from a closed box	Box lid makes a noise on grab and put-down
Grab	Pick-up a bank note from a table	Money makes a noise on grab and put-down
Take (a photo)	Pick-up a camera and point it at a dog	Camera makes a shutter noise when facing dog
Press	Push down on an industrial button	Button compresses when pushed
Pull	Grab rope, pull away from fitting	Rope extends as if pulled out from fitting
Turn on	Push hand into lightswitch	Lightswitch gets depresses, makes clicking noise
Raise	Pick-up an umbrella and hold above head	Raindrops are blocked by umbrella
Brush	Pick-up toothbrush and bring to mouth	Brushing sound is played
Set/place	Pick-up a cup and place on a tray	Cup makes a noise on connection with tray
Cut	Pick-up knife and moved into bread	Slice of bread is cut from loaf, makes noise

own setting. This suggests the participants were experienced in using IVR hardware.

The average age of valid participants was 27 (SD = 6.75). Participant gender skewed heavily male (38) over female (8) or other/did not say (2). Valid participants had a low knowledge of the target learning words during the pre-test, with the average participant knowing less than one word (M = .15; SD = .46).

The majority of valid participants were fluent in more than one language (17 reported as fluent in one language, 26 self-reported as fluent in two languages, 4 in three languages, and one in four languages). We did not find a significant correlation between languages known and learning outcome (r = 0.24, p = 0.10).

Interaction condition was randomly assigned inside the software once it was downloaded onto a participant's computer. As such, 27 participants were assigned to the "action" condition and 21 to the "gesture" condition.

3.4 Corpus

Participants were tested on their knowledge of 15 concrete action verbs. Action verbs were chosen as they are highly embodied and were used in previous gesture and action word memorisation comparisons [58]. The target words were chosen to be familiar actions that allowed for mostly distinct gestures for each word.

Japanese gairaigo (import words) were specifically avoided to reduce the chance of participants' inferring a meaning from their similarity to English. We also attempted to reduce the use of phonetically similar and particularly long Japanese words, as we were concerned that beginner-level learners would find these words difficult to tell apart. A list of these words, the user's action (for the action condition) and the feedback given by the system can be found in Table 1

3.5 Environment

We created an abstract 3D environment in Unity. The environment was explorable via a head-mounted display and embodied controllers. Navigation could be done by moving around the real space and/or by using the thumbsticks on the controllers.

3.6 Evaluation

Participants' knowledge of the verbs was measured in three tests: one administered before their exposure to the environment (pre-test); one immediately after (post-test), and one seven days later (week-test). Participants performed the same test each time, listening to a Japanese word and choosing the English meaning from a list of 15. All three tests were conducted remotely outside of laboratory conditions; the first two were conducted inside IVR and the third was via a web browser. The time taken for each question was timed to help us evaluate the testing sessions. A visual examination of this data did not highlight any individual user taking a consistently long or short time to answer each question, suggesting that participants avoided looking-up answers; being consistently distracted (in a way that could be measured by time) during the evaluation; or rapidly entering answers in order to receive payment. Participants were not given feedback when submitting answers.

Learning gain was calculated as a normalised score between 0 and 1, measured as post-test score minus pre-test score, divided by the number of eligible words for their session. Five participants had existing knowledge of either one or two of the target verbs - these were removed from their pre-test, post-test and eligible words calculations. We tracked whether participants listened to the audio clip before submitting an answer - this was the case for every entry except one, who missed one question. We believe this was the result of an accidental double-input on the previous question, and so removed this question from the participant's score when calculating the normalised result.

After using the system, participants were asked to complete a survey in-browser. This consisted of the Gonzalez-Franco immersive VR embodiment questionnaire [8], the Igroup presence questionnaire [48] and the Intrinsic Motivation Inventory intrinsic motivation questionnaire [42]).

The Gonzalez-Franco embodiment questionnaire was chosen as it is, to our knowledge, the only attempt at a standardised embodiment questionnaire for IVR research. Its division of embodiment into six sub-scales for differing experimental interests allowed us to isolate the two forms of embodiment particularly relevant to this research - agency and body ownership.

The Igroup presence questionnaire was also chosen due to its ability to measure sub-types of presence. We included questions for all four types - general presence (the "general sense of being there"), spatial presence ("the sense of being physically present in the IVR"), involvement ("measuring the attention devoted to the IVR and the involvement experienced"), experienced realism ("measuring the subjective experience of realism in the IVR") [48]. Each of these types could have implications for the cognitive perception of actions and gestures in the IVR. Igroup is also a well-validated method [28][55], and asking participants for their evaluation of presence experienced is considered the most direct way to assess presence [15].

The intrinsic motivation inventory is a well-established tool for measuring sub-scales of motivation [31]. As a learning experience, we determined that the interest/enjoyment and value/usefulness sub-scales should be explored.

3.7 Analysis

We tested our first hypothesis (does the actions group demonstrate stronger verb learning gains than the gesture group) by coding correct responses as 1 and incorrect responses as 0. Where a participant had answered correctly in the pre-test, their future responses for that word were removed.

We used Mixed Models to account for both the fixed (interaction type) and potential random (users, words) effects, as recommended by Macedonia et al. [25]. As the dependent variable was binomial, we used a Generalised Linear Mixed Model.

To test our second hypothesis (the actions group will demonstrate faster response times than the gesture group), we used a Linear Mixed Model due to the continuous dependent variable of response time. Only correct answers were included in the dataset, and outliers were removed. Outliers were highlighted by checking for 1.5 * interquartile range above the third quartile, or below the first quartile. We felt comfortable removing these outliers as they were split fairly evenly between groups, and some participant's mean answer times were skewed by a few longer entries, potentially caused by distracting out-of-lab circumstances.

For our third hypothesis (the actions group will report stronger embodiment than the gesture group), we calculated linear regressions between each of the two embodiment scores calculated from survey results (ownership and agency) and the interaction condition (action or gesture).

For our fourth hypothesis (the actions group will report stronger presence than the gesture group), we calculated linear regressions between each of the four presences scores calculated from survey results (general, spatial, involvement and realism) and the interaction condition (action or gesture).

For our fifth hypothesis (the actions group will report stronger intrinsic motivation than the gesture group), we calculated each of the linear regression motivations scores (interest, value/usefulness) calculated from survey results and the interaction condition (action or gesture).

4 RESULTS

Our comparison of pre-test results of included participants found no significant difference (t = 1.31; p = .20) between the pre-existing knowledge of the action (m = 0.01) and gesture groups (m = 0; 1 being perfect knowledge of all 15 words), with only five participants knowing any Japanese.

4.1 The actions group will demonstrate stronger verb learning gains than the gesture group

The descriptive results for both post-test and one-week learning gain are presented in Table 2, and the GLMM results are presented in Table 3.

For the post-test, our GLMM (n = 720; 48 participants) showed learning gain varied across both participants (σ^2 = 1.88) and words (σ^2 = 0.58) After controlling for these random factors, the model presented a statistically significant relationship between interaction

Table 2: Table of learning gain results from tests immediately after the session (post-test) and one week later (week-test)

Results	N	Mean Score	Mean RT
Action: Post-test	27	0.66 ±0.27	9.13 ±4.84
Gesture: Post-test	21	0.47 ±0.25	8.90 ±3.92
Action: Week-test	21	0.39 ±0.25	5.76 ±4.53
Gesture: Week-test	14	0.27 ±0.19	5.35 ±4.85

Table 3: Table of Generalised Linear Mixed Model results for learning gain. Note: co-efficients are logit

Parameter	Beta	Lower-95	Upper-95	Std. Error
Post-test				
Intercept	-0.13	-0.90	0.64	0.38
Interaction (Action)	1.12	0.24	2.04	0.44
Week-test				
Intercept	-1.30	-2.04	-0.63	3.46
Interaction (Action)	0.79	-0.03	1.64	0.41

type and learning gain (p = .012). Words encoded in the action group were better remembered than those in the gesture group (β = 1.12, 95% CI [0.24,2.04]). In our model, given a participant and word with average intercepts, if they were assigned to the action condition, they would be 26% more likely to correctly remember a word than in the gesture condition (73% vs 47%).

For the one-week follow-up test, our model (n = 525; 36 participants) also showed learning gain varied across both participants (σ^2 = 1.00) and words (σ^2 = 0.22), but likely to a lesser extent. It did not present a significant distinction in learning gain between words encoded in the action group (β = 0.22, 95% CI [-0.03,1.64]) and those in the gesture group. Although not significant, in our model, given a participant and word with average intercepts, the probability of getting a correct response increases from 21% to 37% in the action group.

Therefore h1 is accepted for the immediate post-test results, but not for the week-test.

4.2 The actions group will demonstrate faster response times than the gesture group

The descriptive response time results for both post-test and one-week test are presented in Table 2, and the LMM results for both post-test and one-week later test are presented in Table 4.

For the post-test, our LMM (n = 402; 48 participants) showed response time varied across both participants (σ^2 = 3.66) and words (σ^2 = 3.07). After controlling for these random factors, we were unable to find a significant distinction between between the action group (β = 0.07, 95% CI [-1.35,1.47]) and the gesture group.

For the week-test, our model (n = 173; 33 participants) showed response time varied across words (σ^2 = 4.47) and to a lesser extent users (σ^2 = 0.06). After controlling for these random factors, we were unable to find a significant distinction between between the action group (β = 0.1, 95% CL [-1.28,1.54]) and the gesture group.

Therefore h2 is not accepted for either immediate post-test response times or the week-test.

4.3 Words by InteractionType

After finding repeated evidence of the random effect of words, we used a LMM to explore whether words had an interaction effect with interaction type; to understand if some words were better or less suited to embodied encoding. However, a likelihood ratio test indicated that adding random intercepts for each interaction condition of each word (word*interactionType) did not improve the model over adding random intercepts for each word only. Therefore we cannot conclude that there is a significant interaction between word and interactionType.

Table 4: Table of Linear Mixed Model results for response time

Parameter	Beta	Lower-95	Upper-95	Std. Error
Post-test				
Intercept	9.32	7.92	10.76	0.71
Interaction (Action)	0.07	-1.35	1.47	0.70
Week-test				
Intercept	5.82	4.19	7.48	0.81
Interaction (Action)	0.11	-1.28	1.54	0.70

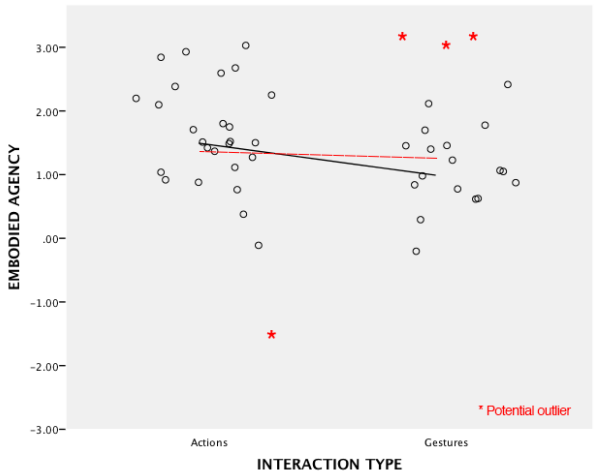


Figure 3: Jittered plot showing relationship between self-reported embodied agency and interaction type. Black line shows significant relationship when outliers removed, red line shows relationship without significance when outliers are included.

4.4 The actions group will report stronger embodiment than the gesture group

Our results did not show a significant correlation between the interaction type (action or gesture) and the self-reported feeling of embodied agency (r = 0.64; s = .67).

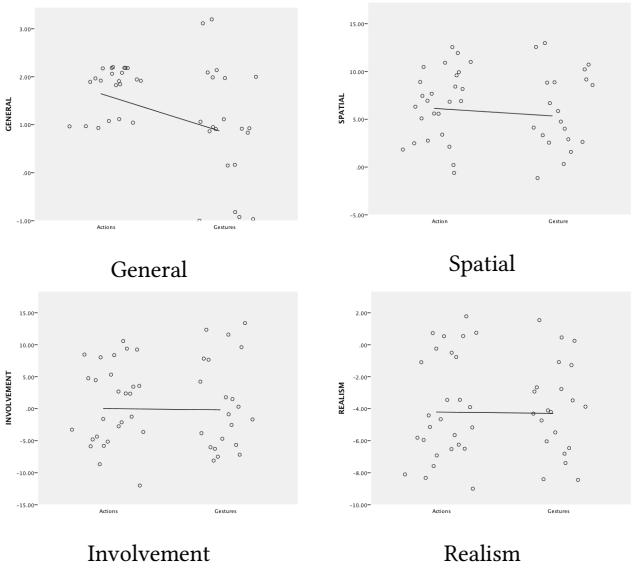
However, we observed four potential outliers (based on inter-quartile range), and when these were removed, our results showed a significant correlation (r = 0.33; p = .03), which would mean interaction type explains 10.8% of the variability of the embodied agency score. A graph depicting the linear correlations between embodied agency and interaction type, both with and without outliers, is presented in Fig. 3.

It was difficult to determine whether these were true outliers or not. These four participants presented embodiment ratings that appear distinct from their peers, however it is not impossible for them to have felt incredibly embodied (or non-embodied) by the interactions, or to have interpreted the question notably differently from others. One of the outlier participants entered universally the lowest scores for all embodied agency questions, but provided more varied results for other questions. Without strong evidence to remove these outliers, however, we have included them in the dataset and so are not able to report a significant relationship.

The relationship with embodied ownership (r = 0.05, p = .71) was not significant.

Table 5: Summary of presence scores and the size and significance of their relationship with interaction type

Presence Type	Action Mean	Gesture Mean	R	R2	P
General	1.7 (±1.1)	1.0 (±1.2)	.32	.103	.026
Spatial	6.0 (±4.3)	5.0 (±5.2)	.01	.012	.045
Involvement	0.6 (±5.9)	0.4 (±6.8)	.02		.091
Realism	-4.0 (±3.2)	-4.0 (±2.9)	.01		.092



General Spatial

Involvement Realism

Figure 4: Graphs depicting linear correlations between interaction type and presence scores, arranged by presence types (general, spatial, involvement and realism). There were significant relationships for general and spatial presence, with a small influence on spatial presence score.

Therefore h3 is not accepted.

4.5 The actions group will report stronger presence than the gesture group

Our results show a significant correlation between the interaction type and self-reported general presence (r = 0.32, p = .026), which means interaction type explains 10.3% of the variability of the general presence score.

Our results also show a significant correlation between the interaction type and spatial presence (r = 0.01, p = .045), which means interaction type explains 1.2% of the variability of the spatial presence score.

We found no significant correlations between interaction type and the involvement presence score (r = 0.016, p = .091) or realism presence score (r = 0.014, p = .092).

A summary of the presence scores is presented in Table 5, and linear correlations for each of these presence measures and the two interaction types is presented in Fig. 4.

Therefore h4 is accepted for general presence and spatial presence, but not for the involvement or realism presence variations.

4.6 The actions group will report stronger motivation than the gesture group

We found no significance in the correlations between interaction type and interest motivation (r = 0.049, p = .741) or value/usefulness motivation (r = 0.013, p = .931).

Therefore h5 is not accepted.

5 DISCUSSION

5.1 Evidence sensorimotor-embodied interactions are actions, not gestures

Our results show that verb learners who take actions on objects in IVR achieved significant and large memorisation gains over learners who make gestures without manipulating objects. This was reflected in immediate learning gain scores, but not by response times. These results have obvious implications for designing optimal IVR-based action-verb learning applications, which should activate users sensorimotor-systems in a form that includes objects for the interaction and feedback from the system for the object's congruent use.

The explanation for why we saw these results, and what that means for sensorimotor-embodied controllers in IVR, is more nuanced. It is possible that the learning gain differences between action and gesture conditions in the IVR can be explained by the same cognitive phenomena that has previously been evidenced in physical world comparative studies between action and gesture. If Wakefield's explanation of different encoding pathways between actions and gestures in the physical world [58] is true, then it is likely that we are seeing a similar results in our IVR.

Extending this further, this means that embodied controllers in IVR provide a cognitive experience similar to that in the physical world - interacting with objects in the physical or virtual worlds are actions; while activating the body to make movements that do not interact with objects, in the physical or virtual worlds, are gestures. Therefore when considering interactive actions in IVR, we should discuss them from the perspective of action-based embodiment theory, rather than gesture-based theory. In short: our interactions in IVR are based upon what we are experiencing in IVR, and not the controllers or physical world bodily movements.

However, there is also another potential explanation for the learning distinctions, which stems from the explanation of the enactment effect as enhancing memory traces [36]. Our results could be highlighting the added learning efficacy that stems from contextually-deployed system feedback and richer situational encoding offered by the action condition. The feedback is two-fold: first from the virtual objects being able to be moved, and second from the system responding to user's manipulations of objects with sound effects or system events. If this was the explanation, we would be unable to extrapolate whether embodied controller users cognitively contextualise their body-based movements as action or

gesture (and even whether that distinction was meaningful, given that we are explaining enhanced learning through memory traces rather than sensorimotor encoding pathways).

These two explanations lead us into an interesting place regarding the simulation of actions in IVR, in that the amount of "memory trace" could be adjusted for exploration in a way not possible in the physical world. For example, in the physical world, it is unlikely that you can separate actions from their contextual environmental feedback - pouring a jug of water will always cause water to fall (unless, perhaps, you're in space). However, in the IVR space, we are able to have different forms of action-feedback that do not correspond to the real world. Whether water falls, how it falls, does not fall, floats upwards or even exists as at all are the choice of the systems' designers.

An interesting additional exploration to provide further clarity on the cause of encoding differences in the IVR would be to amend environmental feedback to gestures (the gesture for pouring water would not move the object, but would play a pouring sound), or stripping environmental feedback from actions (you could move a jug to pour water, but no water or sound runs out from the jug), and contrast these results to our virtual recreations of typical physical world gestures and action processes. We may find that both interactional and environmental feedback are needed for us to contextualise embodied controller actions in IVR in a way similar to physical actions, or that the benefits can be added to gestures through environmental feedback.

5.2 Missing retention, similar response times

We did not find a significant difference in learning retention after one week between the two encoding conditions, although the action condition showed a higher, non-significant, mean learning gain. This is similar to results in previous work [38]. There are three potential explanations for the difference in significance between the immediate and one-week later tests: (1) the drop in participants (from 48 to 35), as many did not complete the one-week later test, reduced the sensitivity of the test; (2) the difference in learning between the conditions is reduced but not eliminated, reducing the sensitivity of the test; (3) learning gain differences between action and gesture only occur immediately, and longer-term learning is similar between conditions. We believe that (1) and (2) are the most likely explanations for our results, as a reduction in the learning difference between experimental groups over time is a pattern familiar in language learning investigations [41][25], and also an artefact of a somewhat artificial language encoding experience.

We found no distinction in response times between the action and gesture conditions. As faster response times are typically associated with stronger encoding [25][7], we would have expected to see faster response times for the action condition to match the learning gain scores. However, given that the difference between the learning gain of the two conditions was so large, and that response time is a less direct measure of learning outcome, we are confident in claiming a distinction between the two conditions.

5.3 Presence, embodiment and motivation metrics

Our results show that participants in the action condition had significantly higher feelings of general presence and spatial presence (albeit with a small correlation), but not involvement or realism. These results suggest that being able to interact with objects in a virtual space enhances the sense of being physically present in the IVR. As the experiment was targeting learning in an abstract space, it could be that no distinctions between involvement or realism were found due the already involved nature of any learning process, or the unrealistic environmental setting.

It was a little surprising to not find any significant relationships between the interaction types and our embodiment measures of ownership or agency. It seems reasonable to assume that of our subjective measures, these would be the most likely to be affected by the different interaction types, as we would expect to see higher levels of self-reported embodiment for the interactive object-manipulation. This result presents questions over the relevancy or efficacy for this embodiment survey [8] for this type of exploration of sensorimotor, interactive embodiment. The agency-related questions in the survey ask about visuo-motor synchronous stimulation (e.g. "It felt like I could control the virtual hand as if it was my own"), which according to our results, appear to be experienced consistently whether you are interacting with virtual objects or not. Perhaps embodiment in a virtual body that can interact with the virtual space is an additional factor that needs a separate survey categorisation.

6 LIMITATIONS

The study participant demographics are a notable limitation of this study, as we used participants who were both familiar with VR (enough to own their own headset) and who had a large enough interest in the technology that they were members of an online community for it. A major implication for this is that the audience might be self-selecting: those who IVR resonate with are potentially more likely to have invested in the hardware than the general populous, and so may be more keenly affected by its affordances. Our sample was also heavily skewed towards men, who have been shown as less likely to suffer from simulator sickness with the current incarnation of IVR technology [52].

There are also limitations to the generalisability of this research to other uses of IVR. For example, it is not clear if the evidence presented here for the similarity of benefits between action-based learning in IVR and in the physical world, would work for other academic subjects (e.g. mathematics) or other areas (e.g. empathy-training, rather than cognitive learning).

Finally, this study uses highly sensorimotor-embodied words: concrete action verbs. Further study of words more peripherally linked to actions, such as nouns, adjectives, and abstract or stative verbs are needed - although this is also true outside of IVR investigations (existing research suggests that while "the sensorimotor neural network is engaged in both concrete and abstract language contents ... concrete multi-word processing relies more on the sensorimotor system, and abstract multi-word processing relies more on the linguistic system" [43]).

7 CONCLUSION

Our findings show that users can have distinct learning outcomes from embodied controller-enabled interactions in IVR based on whether those interactions were presented as actions (i.e. were able to interact with objects) or gestures (i.e. were not). Learners who encoded information while doing actions had significantly better learning outcomes than those who encoded with gestures.

This result is similar to action vs. gesture comparisons conducted in the physical world. If we subscribe to the view that humans memorise information differently depending on whether it was encoded using an actions or gesture, these results could mean that participants had cognitive experiences in IVR that were similar to physical world actions and gesture experiences. This suggests that our cognitive perceptions of interactions in IVR are not restricted by the controllers or abstracted physical world bodily movements, but by what we are experiencing inside IVR.

While this has only been evidenced for memorization (in this study), if this were the case generally, it would mean that we should consider sensorimotor actions taken in IVR in the same way we contextualise actions in the physical world, and this could have implications for the emerging use of IVR in PTSD or exposure therapy. We hope these results will encourage further study in these other areas.

However, the observed learning difference could also be explained by theories around encoding depth, and that our actions in IVR provided additional interactive feedback, which the gestures did not. If this was the case, then it is more difficult to outline a strong case for how we cognitively contextualise our interactions with embodied controllers and IVR. Further research is needed to determine which of these explanations might be the case.

REFERENCES

[1] Michael L Anderson, Michael J Richardson, and Anthony Chemero. 2012. Eroding the boundaries of cognition: Implications of embodiment 1. *Topics in cognitive science* 4, 4 (2012), 717–730.

[2] Shannon KT Bailey, Cheryl I Johnson, and Valerie K Sims. 2018. Using Natural Gesture Interactions Leads to Higher Usability and Presence in a Computer Lesson. In *Congress of the International Ergonomics Association*. Springer, 663–671.

[3] Domna Banakou, Sameer Kishore, and Mel Slater. 2018. Virtually being einstein results in an improvement in cognitive task performance and a decrease in age bias. *Frontiers in psychology* 9 (2018), 917.

[4] Ronald L Cohen. 1989. Memory for action events: The power of enactment. *Educational psychology review* 1, 1 (1989), 57–80.

[5] Johannes Engelkamp. 1998. *Memory for actions*. Psychology Press/Taylor & Francis (UK).

[6] Johannes Engelkamp and Horst Krumnacker. 1980. Image-and motor-processes in the retention of verbal materials. *Zeitschrift für experimentelle und angewandte Psychologie* (1980).

[7] Orly Fuhrman, Anabel Eckerling, Naama Friedmann, Ricardo Tarrasch, and Gal Raz. 2020. The moving learner: Object manipulation in virtual reality improves vocabulary learning. *Journal of Computer Assisted Learning* (2020).

[8] Mar Gonzalez-Franco and Tabitha C Peck. 2018. Avatar embodiment. towards a standardized questionnaire. *Frontiers in Robotics and AI* 5 (2018), 74.

[9] Rob Gray. 2017. Transfer of training from virtual to real baseball batting. *Frontiers in psychology* 8 (2017), 2183.

[10] Jérôme Guegan, Stéphanie Buisine, Fabrice Mantelet, Nicolas Maranzana, and Frédéric Segonds. 2016. Avatar-mediated creativity: When embodying inventors makes engineers more creative. *Computers in Human Behavior* 61 (2016), 165–175.

[11] Jeng Hong Ho, Steven ZhiYing Zhou, Dong Wei, and Alfred Low. 2009. Investigating the effects of educational Game with Wii Remote on outcomes of learning. In *Transactions on Edutainment III*. Springer, 240–252.

[12] Susan L Hornstein and Neil W Mulligan. 2001. Memory of action events: The role of objects in memory of self-and other-performed tasks. *The American journal of psychology* 114, 2 (2001), 199.

[13] Autumn B Hostetter and Martha W Alibali. 2008. Visible embodiment: Gestures as simulated action. *Psychonomic bulletin & review* 15, 3 (2008), 495–514.

[14] Autumn B Hostetter, Wim Pouw, and Elizabeth M Wakefield. 2020. Learning from gesture and action: An investigation of memory for where objects went and how they got there. *Cognitive Science* 44, 9 (2020), e12889.

[15] Wijnand A IJsselsteijn, Huib De Ridder, Jonathan Freeman, and Steve E Avons. 2000. Presence: concept, determinants, and measurement. In *Human vision and electronic imaging V*, Vol. 3959. International Society for Optics and Photonics, 520–529.

[16] Spencer Kelly, Meghan Healey, Asli Özyürek, and Judith Holler. 2015. The processing of speech, gesture, and action during language comprehension. *Psychonomic bulletin & review* 22, 2 (2015), 517–523.

[17] Konstantina Kilteni, Ilias Bergstrom, and Mel Slater. 2013. Drumming in immersive virtual reality: the body shapes the way we play. *IEEE transactions on visualization and computer graphics* 19, 4 (2013), 597–605.

[18] Reza Kormi-Nouri and Lars-Göran Nilsson. 2001. The Motor Component. *Memory for action: A distinct form of episodic memory?* (2001), 97.

[19] Rotem Lammfromm and Daniel Gopher. 2011. Transfer of skill from a virtual reality trainer to real juggling. In *BIO web of conferences*, Vol. 1. EDP Sciences, 00054.

[20] Elinda Ai-Lim Lee, Kok Wai Wong, and Chun Che Fung. 2010. How does desktop virtual reality enhance learning outcomes? A structural equation modeling approach. *Computers & Education* 55, 4 (2010), 1424–1442.

[21] Seokwon Lee, Kihong Park, Junyeop Lee, and Kibum Kim. 2017. User study of VR basic controller and data glove as hand gesture inputs in VR games. In *2017 international symposium on ubiquitous virtual reality (isuvr)*. IEEE, 1–3.

[22] Wan-Ju Lee, Chi-Wen Huang, Chia-Jung Wu, Shing-Tsaan Huang, and Gwo-Dong Chen. 2012. The effects of using embodied interactions to improve learning performance. In *2012 IEEE 12th International Conference on Advanced Learning Technologies*. IEEE, 557–559.

[23] Gongfa Li, Heng Tang, Ying Sun, Jianyi Kong, Guozhang Jiang, Du Jiang, Bo Tao, Shuang Xu, and Honghai Liu. 2019. Hand gesture recognition based on convolution neural network. *Cluster Computing* 22, 2 (2019), 2719–2729.

[24] Chien-Yu Lin, Yen-Huai Jen, Li-Chih Wang, Ho-Hsiu Lin, and Ling-Wei Chang. 2011. Assessment of the application of Wii remote for the design of interactive teaching materials. In *International Conference on Information and Management Engineering*. Springer, 483–490.

[25] Manuela Macedonia, AE Lehner, and C Repetto. 2020. Positive effects of grasping virtual objects on memory for novel words in a second language. *Scientific Reports* 10, 1 (2020), 1–13.

[26] Colin M MacLeod, Nigel Gopie, Kathleen L Hourihan, Karen R Neary, and Jason D Ozubko. 2010. The production effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 36, 3 (2010), 671.

[27] Christopher R Madan and Anthony Singhal. 2012. Encoding the world around us: Motor-related processing influences verbal memory. *Consciousness and Cognition* 21, 3 (2012), 1563–1570.

[28] Guido Makransky, Lau Lilleholt, and Anders Aaby. 2017. Development and validation of the Multimodal Presence Scale for virtual reality environments: A confirmatory factor analysis and item response theory approach. *Computers in Human Behavior* 72 (2017), 276–285.

[29] Jennifer A Mangels and Aileen Heinberg. 2006. Improved episodic integration through enactment: Implications for aging. *The Journal of General Psychology* 133, 1 (2006), 37–65.

[30] Taylor Martin and Daniel L Schwartz. 2005. Physically distributed learning: Adapting and reinterpreting physical environments in the development of fraction concepts. *Cognitive science* 29, 4 (2005), 587–625.

[31] Edward McAuley, Terry Duncan, and Vance V Tammen. 1989. Psychometric properties of the Intrinsic Motivation Inventory in a competitive sport setting: A confirmatory factor analysis. *Research quarterly for exercise and sport* 60, 1 (1989), 48–58.

[32] Rory McGloin, Kirstie M Farrar, and Marina Krcmar. 2011. The impact of controller naturalness on spatial presence, gamer enjoyment, and perceived realism in a tennis simulation video game. *Presence: Teleoperators and Virtual Environments* 20, 4 (2011), 309–324.

[33] Tassos A Mikropoulos and Antonis Natsis. 2011. Educational virtual environments: A ten-year review of empirical research (1999–2009). *Computers & Education* 56, 3 (2011), 769–780.

[34] Katherine H Mumford and Sotaro Kita. 2014. Children use gesture to interpret novel verb meanings. *Child Development* 85, 3 (2014), 1181–1189.

[35] Miriam A Novack, Elizabeth M Wakefield, and Susan Goldin-Meadow. 2016. What makes a movement a gesture? *Cognition* 146 (2016), 339–348.

[36] Lars Nyberg. 1993. *The enactment effect: Studies of a memory phenomenon*. Ph.D. Dissertation. Umeå Universitet.

[37] Lars Nyberg, Lars-Göran Nilsson, and Lars Bäckman. 1991. A component analysis of action events. *Psychological Research* 53, 3 (1991), 219–225.

[38] Jack Ratcliffe and Laurissa Tokarchuk. 2020. Evidence for embodied cognition in immersive virtual environments using a second language learning environment. In *2020 IEEE Conference on Games (CoG)*. IEEE, 471–478.

[39] Jack Ratcliffe and Laurissa Tokarchuk. 2020. Presence, embodied interaction and motivation: distinct learning phenomena in an immersive virtual environment. In *Proceedings of the 28th ACM International Conference on Multimedia*. 3661–3668.

[40] Matthias Rauterberg. 1999. From gesture to action: Natural user interfaces. *Technical University of Eindhoven, Mens-Machine Interactive, Diesrede* (1999), 15–25.

[41] Susanne Rott. 1999. THE EFFECT OF EXPOSURE FREQUENCY ON INTERMEDIATE LANGUAGE LEARNERS'INCIDENTAL VOCABULARY ACQUISITION AND RETENTION THROUGH READING. *Studies in second language acquisition* 21, 4 (1999), 589–619.

[42] Richard M Ryan. 1982. Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of personality and social psychology* 43, 3 (1982), 450.

[43] Katrin Sakreida, Claudia Scorolli, Mareike M Menz, Stefan Heim, Anna M Borghi, and Ferdinand Binkofski. 2013. Are abstract action words embodied? An fMRI investigation at the interface between language and motor cognition. *Frontiers in human neuroscience* 7 (2013), 125.

[44] Eli Saltz and Suzanne Donnenwerth-Nolan. 1981. Does motoric imagery facilitate memory for sentences? A selective interference test. *Journal of Verbal Learning and Verbal Behavior* 20, 3 (1981), 322–332.

[45] Marilyn C Salzman, Chris Dede, R Bowen Loftin, and Jim Chen. 1999. A model for understanding how virtual reality aids complex conceptual learning. *Presence: Teleoperators & Virtual Environments* 8, 3 (1999), 293–316.

[46] Emilia Scalona, Juri Taborri, Darren Richard Hayes, Zaccaria Del Prete, Stefano Rossi, and Eduardo Palermo. 2019. Is the Neuromuscular Organization of Throwing Unchanged in Virtual Reality? Implications for Upper Limb Rehabilitation. *Electronics* 8, 12 (2019), 1495.

[47] Mike Schmierbach, Anthony M Limperos, and Julia K Woolley. 2012. Feeling the need for (personalized) speed: How natural controls and customization contribute to enjoyment of a racing game through enhanced immersion. *Cyberpsychology, Behavior, and Social Networking* 15, 7 (2012), 364–369.

[48] Thomas W Schubert. 2003. The sense of presence in virtual environments: A three-component scale measuring spatial presence, involvement, and realness. *Z. für Medienpsychologie* 15, 2 (2003), 69–71.

[49] Moamer Shakroum, Kok Wai Wong, and Chun Che Fung. 2018. The influence of gesture-based learning system (GBLS) on learning outcomes. *Computers & Education* 117 (2018), 75–101.

[50] David M Sidhu and Penny M Pexman. 2016. Is moving more memorable than proving? Effects of embodiment and imagined enactment on verb memory. *Frontiers in psychology* 7 (2016), 1010.

[51] Mel Slater, Bernhard Spanlang, Maria V Sanchez-Vives, and Olaf Blanke. 2010. First person experience of body transfer in virtual reality. *PloS one* 5, 5 (2010), e10564.

[52] Kay Stanney, Cali Fidopiastis, and Linda Foster. 2020. Virtual reality is sexist: but it does not have to be. *Frontiers in Robotics and AI* 7 (2020), 4.

[53] Haichun Sun and Yong Gao. 2016. Impact of an active educational video game on children's motivation, science knowledge, and physical activity. *Journal of Sport and Health Science* 5, 2 (2016), 239–245.

[54] Judith Tirp, Christina Steingröver, Nick Wattie, Joseph Baker, and Jörg Schorer. 2015. Virtual realities as optimal learning environments in sport-A transfer study of virtual and real dart throwing. *Psychological Test and Assessment Modeling* 57, 1 (2015), 57.

[55] Jacinto Vasconcelos-Raposo, Maximino Bessa, Miguel Melo, Luis Barbosa, Rui Rodrigues, Carla Maria Teixeira, Luciana Cabral, and António Augusto Sousa. 2016. Adaptation and validation of the Igroup presence questionnaire (IPQ) in a Portuguese sample. *Presence: Teleoperators and virtual environments* 25, 3 (2016), 191–203.

[56] Christian Vázquez, Lei Xia, Takako Aikawa, and Pattie Maes. 2018. Words in motion: Kinesthetic language learning in virtual reality. In *2018 IEEE 18th International Conference on advanced learning technologies (ICALT)*. IEEE, 272–276.

[57] Elizabeth M Wakefield, Casey Hall, Karin H James, and Susan Goldin-Meadow. 2017. Representational Gesture as a Tool for Promoting Verb Learning in Young Children. Boston University Conference on Language Development, Boston, MA.

[58] Elizabeth M Wakefield, Casey Hall, Karin H James, and Susan Goldin-Meadow. 2018. Gesture for generalization: gesture facilitates flexible learning of words for actions on objects. *Developmental science* 21, 5 (2018), e12656.

[59] Tim Zindulka, Myroslav Bachynskyi, and Jörg Müller. 2020. Performance and Experience of Throwing in Virtual Reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–8.